

Nine Projects, One Codebase

A Static Search Engine for Digital Editions

Martin Holmes

(Humanities Computing and Media Centre at the University of Victoria)

Joey Takeda

(Digital Humanities Innovation Lab at Simon Fraser University)

Project Endings

<https://endings.uvic.ca/>

Project Endings

- Static Websites
 - No server side dependencies
 - Strictly XHTML5 + CSS + Vanilla JS
 - Best chance for preservation 50 years + (we hope)

Endings Projects

- Robert Graves Diary (graves.uvic.ca)
- Map of Early Modern London (mapoflondon.uvic.ca)
- The Winnifred Eaton Archive (winnifredeatonarchive.org)
- Mapping Keats's Progress (johnkeats.uvic.ca)
- My Norse Digital Image Repository (myndir.uvic.ca)
- Le mariage sous L'Ancien Régime (mariage.uvic.ca)
- Digital Victorian Periodical Poetry (dvpp.uvic.ca)
- Landscapes of Injustice (loi.uvic.ca)
- Linked Early Modern Drama Online (lemdo.uvic.ca)

But...

But...

- All of these projects required robust searching mechanisms, including:

But...

- All of these projects required robust searching mechanisms, including:
 - Keyword search

But...

- All of these projects required robust searching mechanisms, including:
 - Keyword search
 - Exact phrase search

But...

- All of these projects required robust searching mechanisms, including:
 - Keyword search
 - Exact phrase search
 - Wildcard search

But...

- All of these projects required robust searching mechanisms, including:
 - Keyword search
 - Exact phrase search
 - Wildcard search
 - Filtered search (by date, by author, +++)

Why not use...?

Why not use...?



Google Search

ENHANCED BY Google



Why not use...?

- Not particularly reliable



unprofitably site:dvpp.uvic.ca



All



Maps



Images



Videos



News



More

Settings

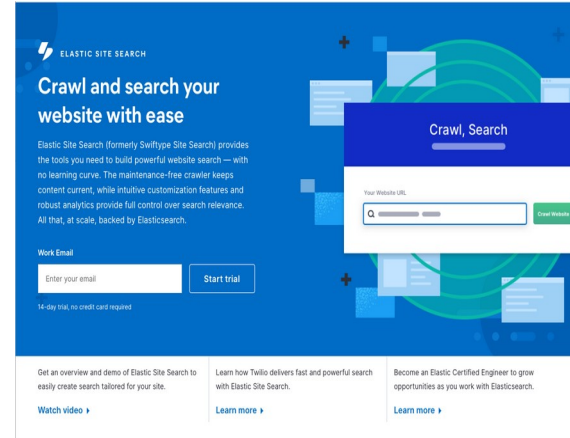
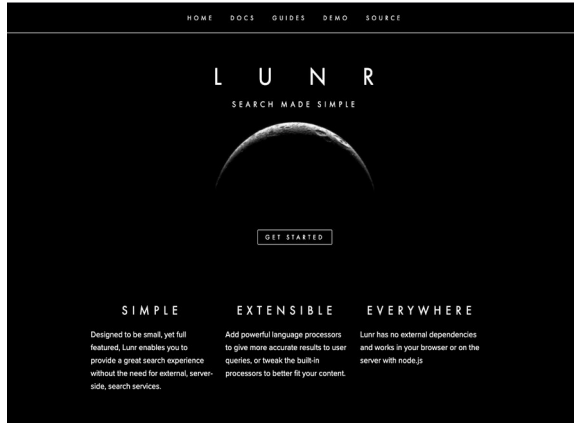
Tools

Your search - **unprofitably site:dvpp.uvic.ca** - did not match any documents.

Suggestions:

- Make sure that all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

Why not use...?



Why not use...?

- Not particularly reliable
- External dependency = technical debt

Why not use...?

- Not particularly reliable
- External dependency = technical debt
- Too many documents to put in one index

So we built our own

projectEndings / staticSearch

Unwatch 3 Unstar 7 Fork 1

Code Issues 10 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

A codebase to support a pure JSON search engine requiring no backend for any XHTML5 document collection

search javascript offline xsl porter-stemmer xhtml5 Manage topics

455 commits 3 branches 0 packages 3 releases 2 contributors View license

Branch: master New pull request Create new file Upload files Find file Clone or download

This branch is 6 commits behind dev. Pull request Compare

martindholmes	Add heading and id to get schemaSpec into documentation.	Latest commit 76b3ae9 9 days ago
docs	Add heading and id to get schemaSpec into documentation.	9 days ago
emod	Laying groundwork for #35. Adding mayoral shows (and the process by w...	3 months ago
jenkins	Add routine warning message to build log parser.	5 months ago
js	Fix for issue 50.	last month
lib	Switching from Saxon 9 to Saxon 10.	2 months ago
schema	Add heading and id to get schemaSpec into documentation.	9 days ago
test	New intro section for documentation.	9 days ago
utilities	Laying groundwork for #35. Adding mayoral shows (and the process by w...	3 months ago
xsl	Tweak to doc processing to fix html title.	9 days ago
.gitignore	Implemented issue #28: an optional setting pointing to a version file...	4 months ago
LICENSE	Initial commit	11 months ago
README.md	Update readme prior to release.	2 months ago
build.xml	Adding documentation to the build and cleaning up the comments	13 days ago
buildSchema.xml	Move schema processing to Saxon 10.	2 months ago
configTest.xml	Implemented item 2 of issue #21: output folder is now configurable in...	4 months ago
configTestWorthReading.xml	Adding a new config for testing #21	3 months ago
config_emod.xml	Laying groundwork for #35. Adding mayoral shows (and the process by w...	3 months ago
config_moeml.xml	Modifying config files to validate against schema after the removal o...	6 months ago
license_BSD.txt	Adding a BSD license	11 months ago
makeRelease.xml	Release build file tested and working.	4 months ago
staticSearch.xpr	Adding more documentation; adding schema build scenario for Oxygen.	4 months ago

README.md

staticSearch

A codebase to support a pure JSON search engine requiring no backend for any XHTML5 document collection

So we built our own

- <https://github.com/projectEndings/staticSearch>

The screenshot shows the GitHub repository page for `projectEndings / staticSearch`. The repository is described as "A codebase to support a pure JSON search engine requiring no backend for any XHTML5 document collection". It has 455 commits, 3 branches, 0 packages, 3 releases, and 2 contributors. The repository is currently on the `master` branch, which is 6 commits behind `dev`. The latest commit by `martindholmes` is titled "Add heading and id to get schemaSpec into documentation." and was made 9 days ago. The repository contains several files and directories, including `docs`, `emod`, `jenkins`, `js`, `lib`, `schema`, `test`, `utilities`, `xsl`, `.gitignore`, `LICENSE`, `README.md`, `build.xml`, `buildSchema.xml`, `configTest.xml`, `configTestWorthReading.xml`, `config_emod.xml`, `config_moeml.xml`, `license_BSD.txt`, `makeRelease.xml`, and `staticSearch.xpr`. The `README.md` file is currently selected and displayed at the bottom of the page.

projectEndings / staticSearch

Unwatch 3 Unstar 7 Fork 1

Code Issues 19 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

A codebase to support a pure JSON search engine requiring no backend for any XHTML5 document collection

search javascript offline xslt porter-stemmer xhtml5 Manage topics

455 commits 3 branches 0 packages 3 releases 2 contributors View license

Branch: master New pull request Create new file Upload files Find file Clone or download +

This branch is 6 commits behind dev. Pull request Compare

martindholmes Add heading and id to get schemaSpec into documentation. Latest commit 76b3ae9 9 days ago

docs	Add heading and id to get schemaSpec into documentation.	9 days ago
emod	Laying groundwork for #35. Adding mayoral shows (and the process by w...	3 months ago
jenkins	Add routine warning message to build log parser.	5 months ago
js	Fix for issue 50.	last month
lib	Switching from Saxon 9 to Saxon 10.	2 months ago
schema	Add heading and id to get schemaSpec into documentation.	9 days ago
test	New intro section for documentation.	9 days ago
utilities	Laying groundwork for #35. Adding mayoral shows (and the process by w...	3 months ago
xsl	Tweak to doc processing to fix html title.	9 days ago
.gitignore	Implemented issue #28: an optional setting pointing to a version file...	4 months ago
LICENSE	Initial commit	11 months ago
README.md	Update readme prior to release.	2 months ago
build.xml	Adding documentation to the build and cleaning up the comments	13 days ago
buildSchema.xml	Move schema processing to Saxon 10.	2 months ago
configTest.xml	Implemented item 2 of issue #21: output folder is now configurable in...	4 months ago
configTestWorthReading.xml	Adding a new config for testing #21	3 months ago
config_emod.xml	Laying groundwork for #35. Adding mayoral shows (and the process by w...	3 months ago
config_moeml.xml	Modifying config files to validate against schema after the removal o...	6 months ago
license_BSD.txt	Adding a BSD license	11 months ago
makeRelease.xml	Release build file tested and working.	4 months ago
staticSearch.xpr	Adding more documentation; adding schema build scenario for Oxygen.	4 months ago

README.md

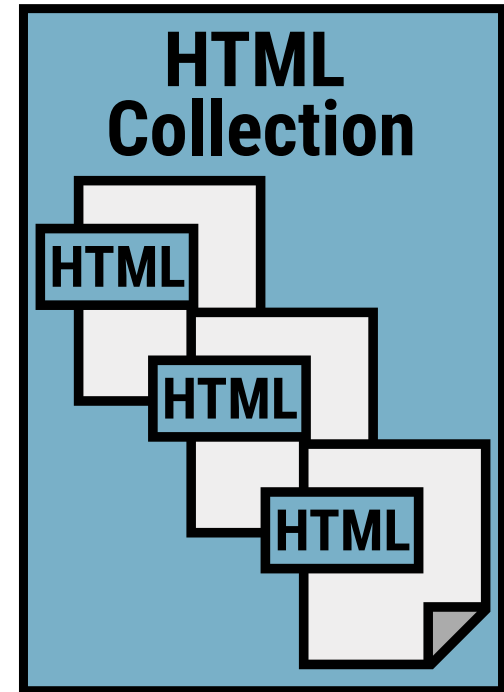
staticSearch

A codebase to support a pure JSON search engine requiring no backend for any XHTML5 document collection

How it works

You have:

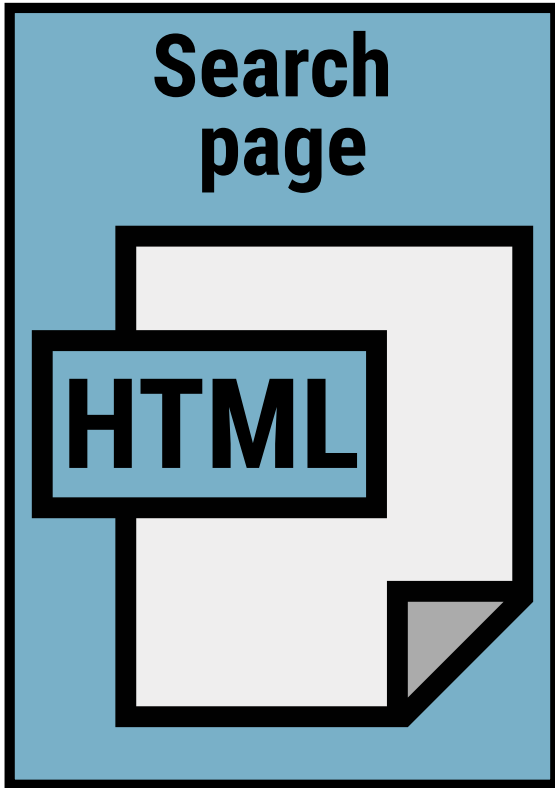
- A collection of HTML files
- All well-formed & valid XHTML



You add:

- An HTML search page containing a special div:

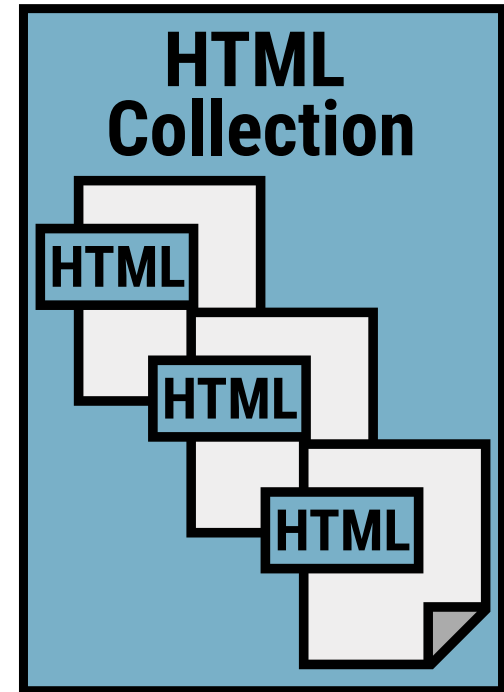
```
<div id="staticSearch"></div>
```



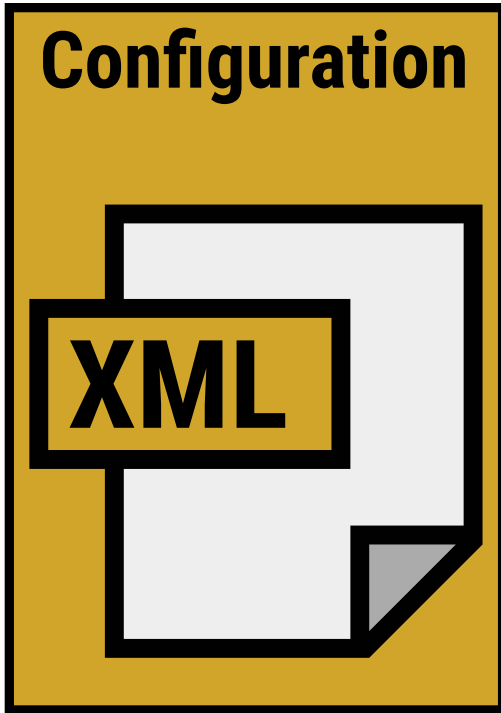
You add:

- Some metadata elements to support search filters:

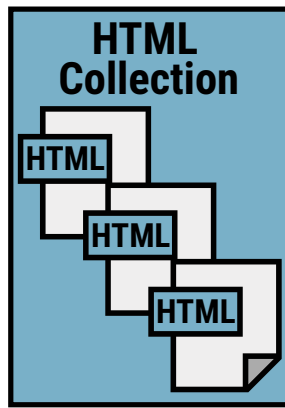
```
<meta name="Document type"  
class="staticSearch.desc"  
content="Poems" />
```

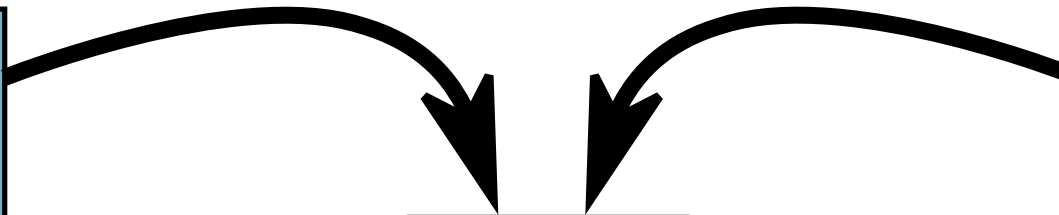
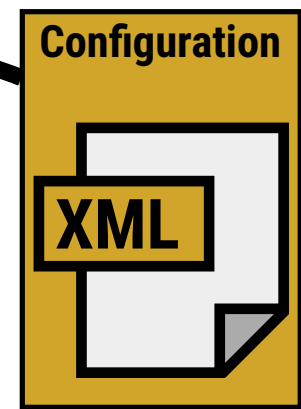
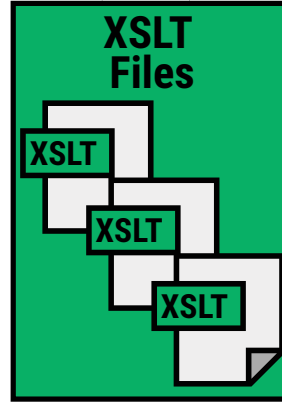
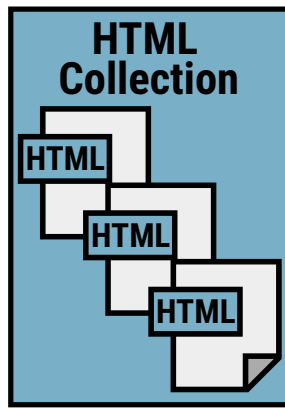


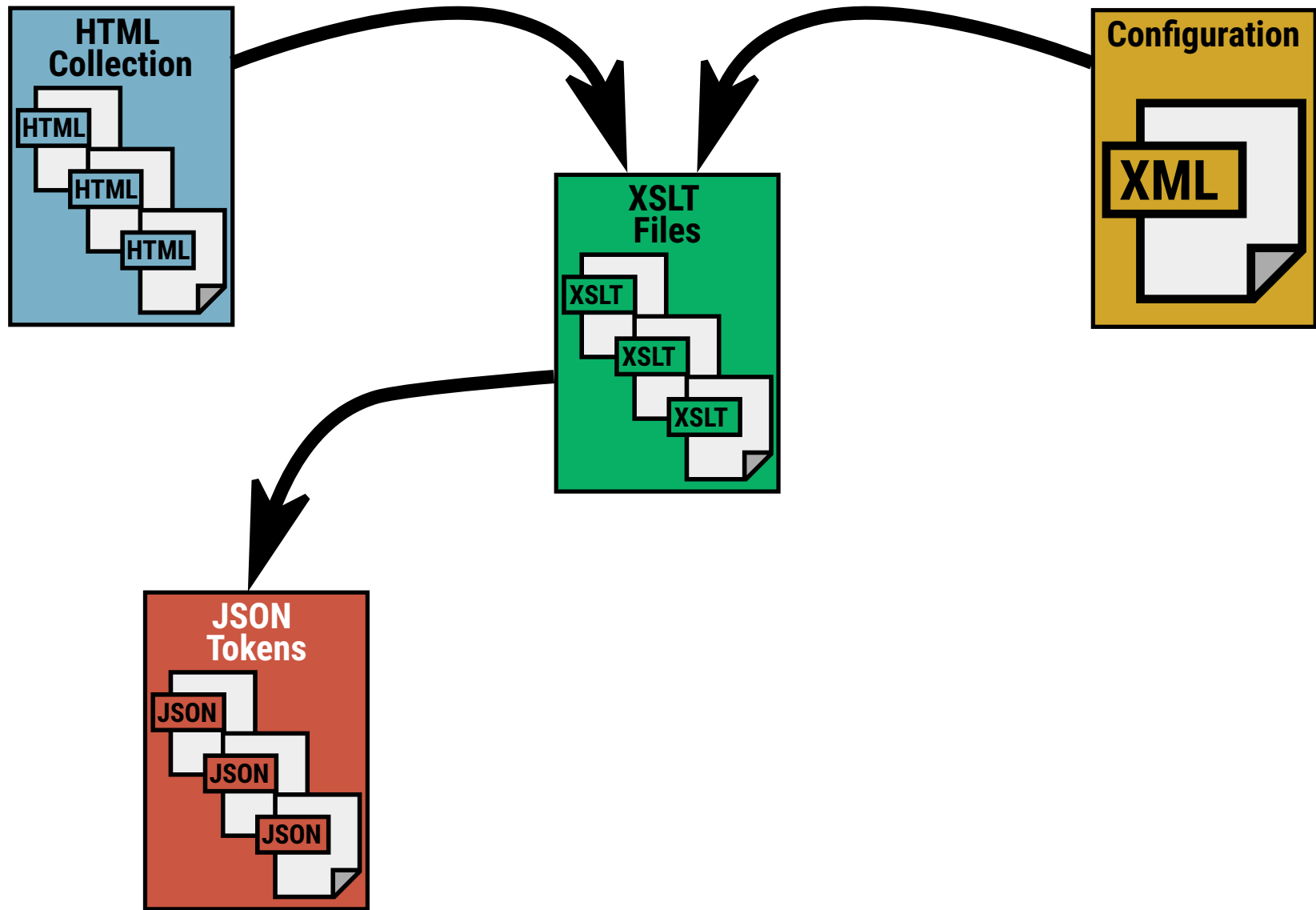
You create:

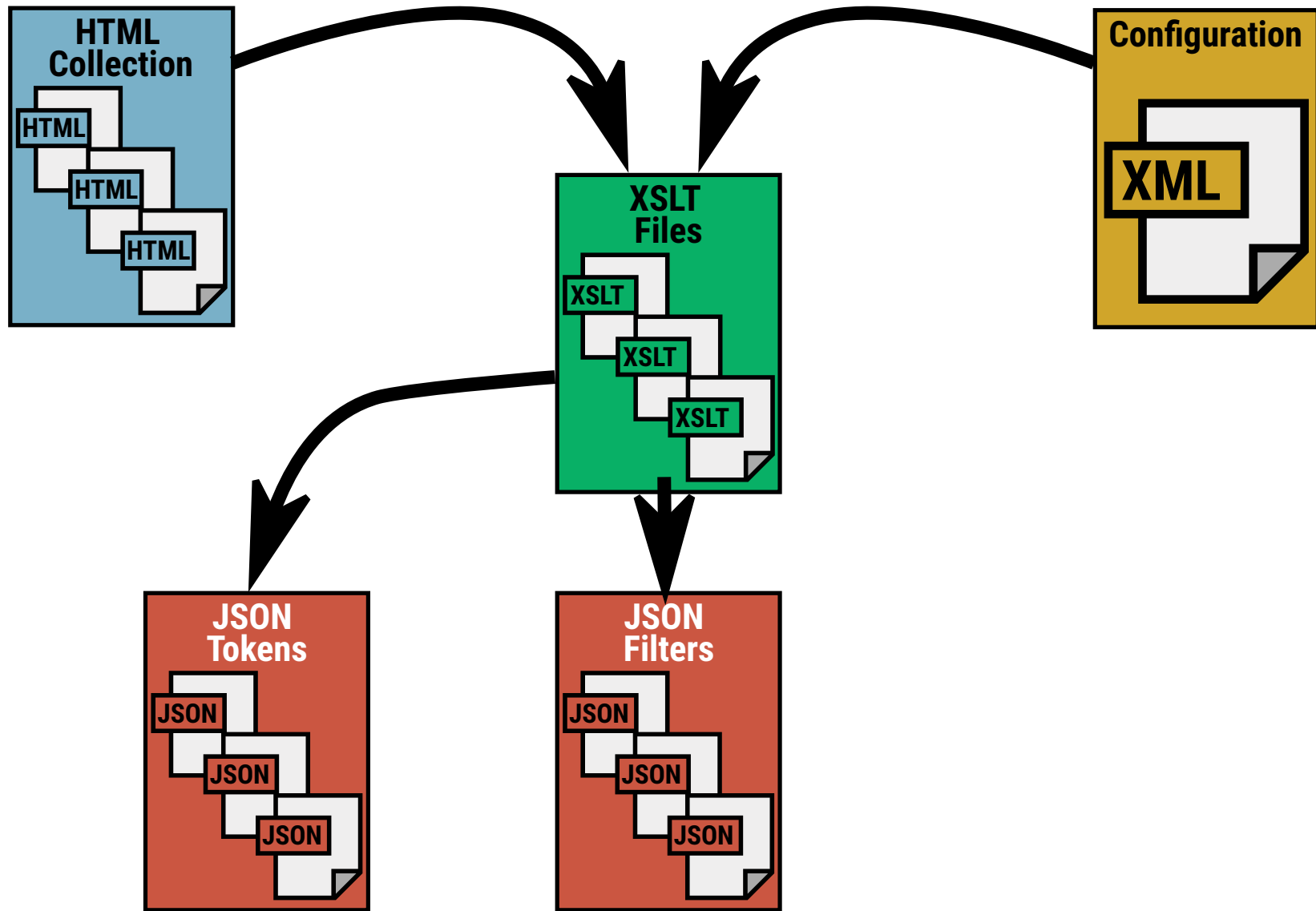


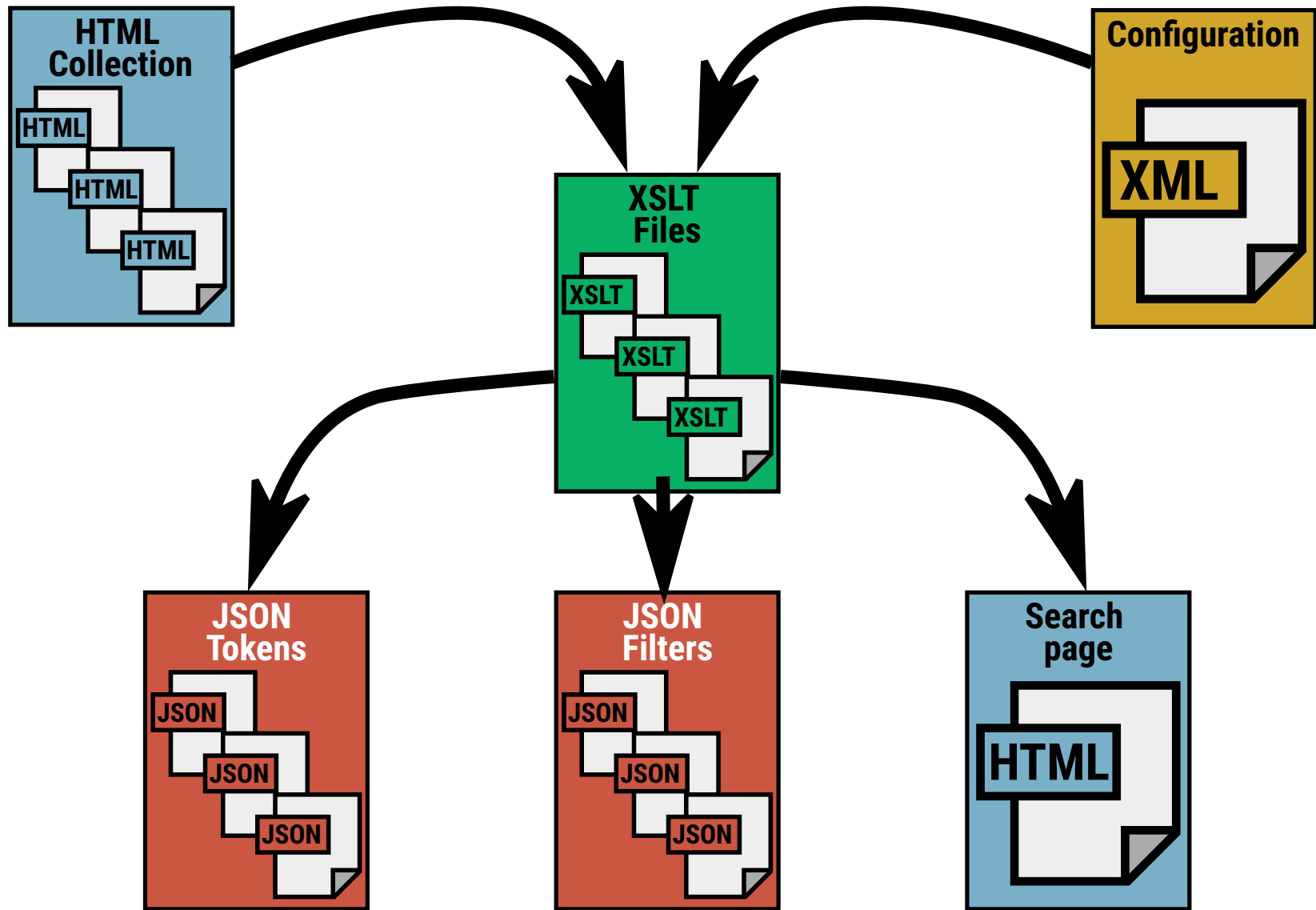
- A configuration file (XML) to specify the features and constraints for your search engine











JSON token file

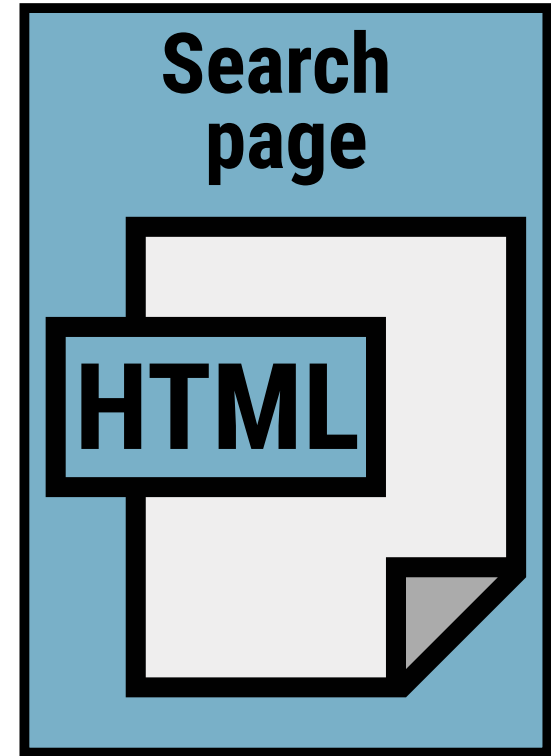
```
{
  "token": "unprofit",
  "instances": [
    {
      "docId": "pom_2025_ithe_old_man_of_hoy",
      "docUri": "poems/goodwords/1870/pom_2025_ithe_old_man_of_hoy.html",
      "score": 1,
      "contexts": [
        {
          "form": "unprofitably",
          "context": "...whole that day was spent <mark>unprofitably</mark>.",
          "weight": 1,
          "pos": 400
        }
      ]
    },
    {
      "docId": "pom_8733_john_and_joan_canto_ii",
      "docUri": "poems/blackwoods/1820/pom_8733_john_and_joan_canto_ii.html",
      "score": 1,
      "contexts": [
        {
          "form": "unprofitable",
          "context": "...too much ap- propriated unto <mark>unprofitable</mark> jocularities and facetiousness. Craving licence,...",
          "weight": 1,
          "pos": 164
        }
      ]
    }
  ]
}
```

Filter JSON

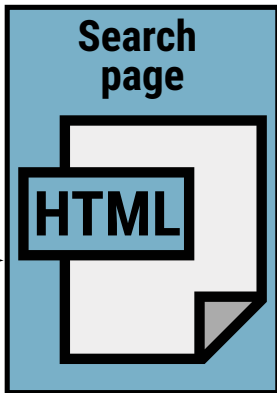
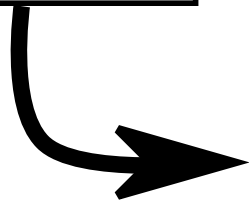
```
{
  "filterId": "ssBool2",
  "filterName": "Unsigned",
  "ssBool2_1": {
    "value": "true",
    "docs": [
      "poems/chambers_series/1867/pom_7594_the_husbands_request.html",
      "poems/alltheyearround/1875/pom_4271_the_hourglass.html",
      "poems/alltheyearround/1879/pom_4507_in_the_conservatory.html",
      "poems/blackwoods/1843/pom_9963_jolly_father_joe.html",
      "poems/blackwoods/1829/pom_10314_the_watchmans_lament.html"
    ]
  }
}
```

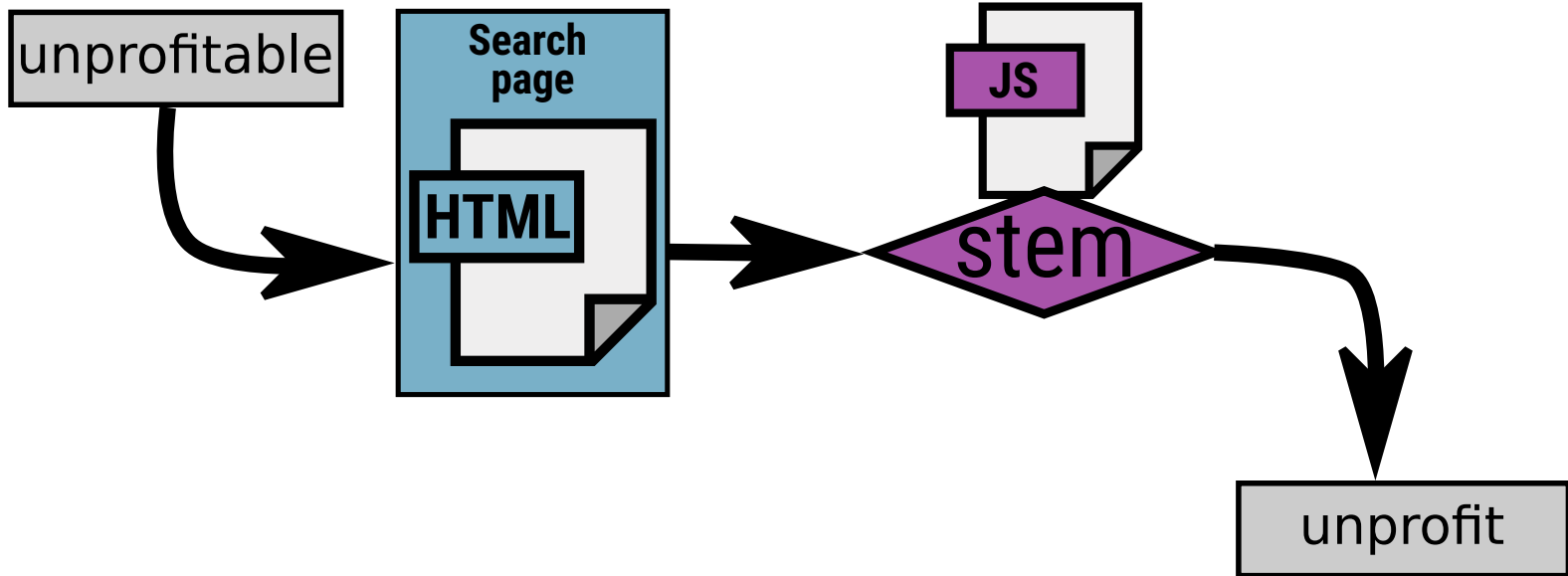
On the website...

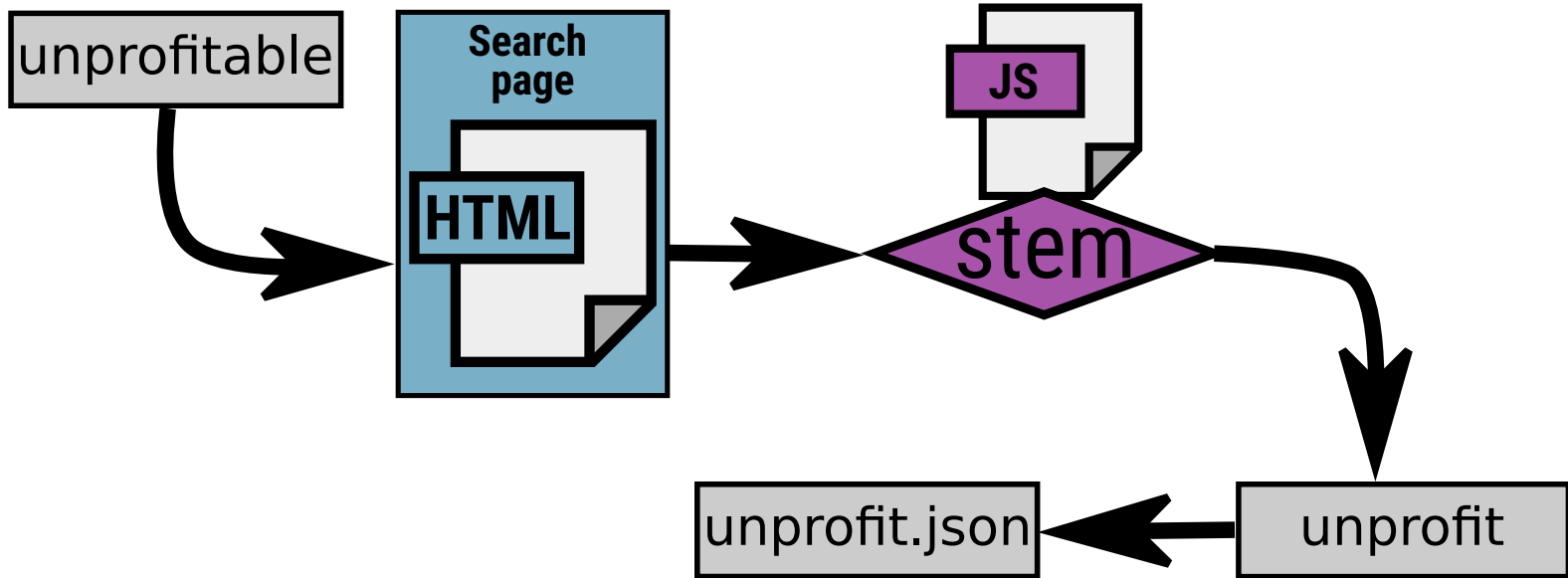
- ...of the thousands of JSON files...
- ...the search page JS retrieves only the ones it needs for the search you do:

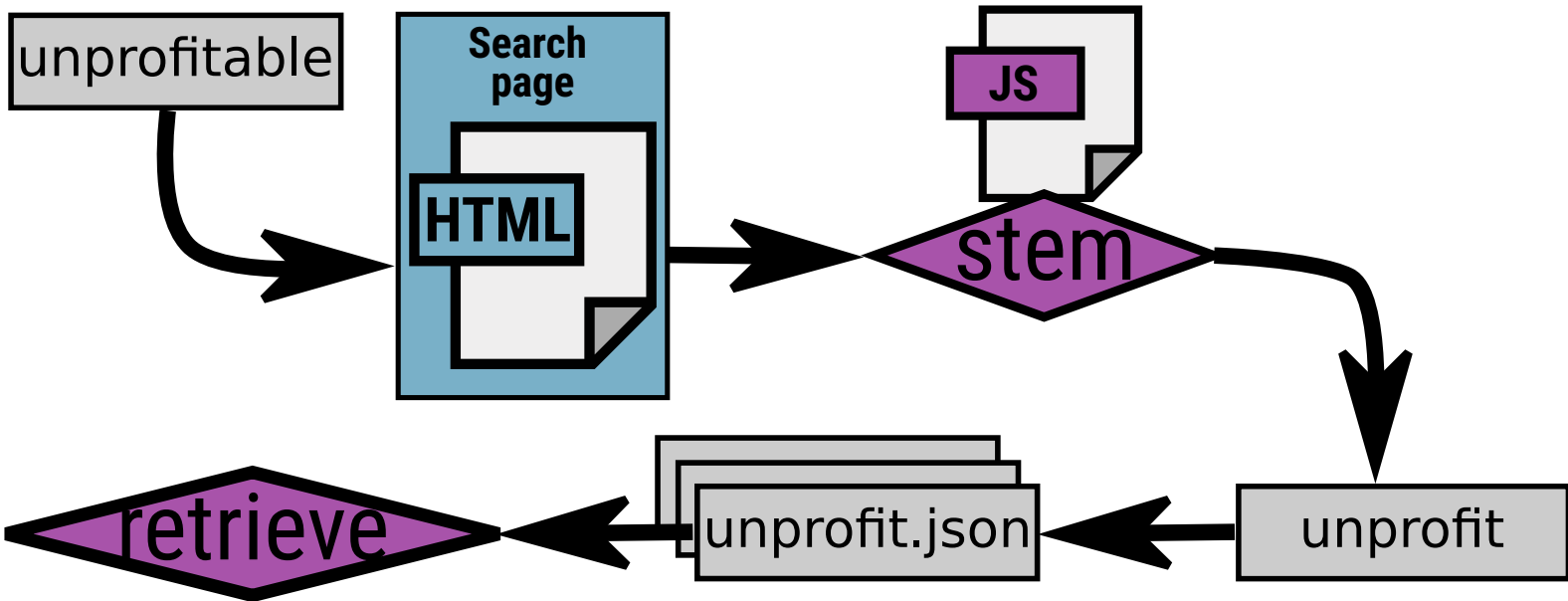


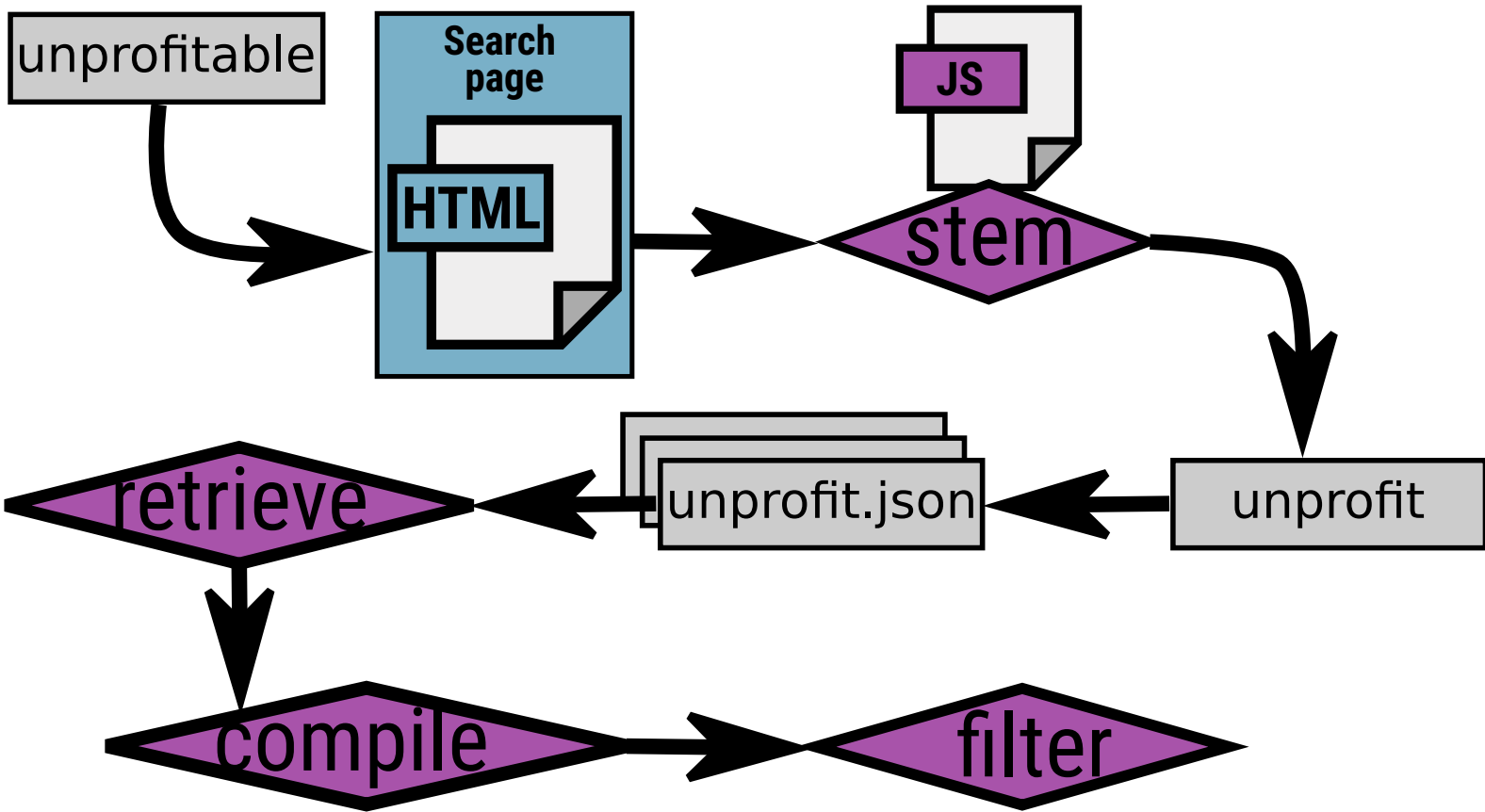
unprofitable

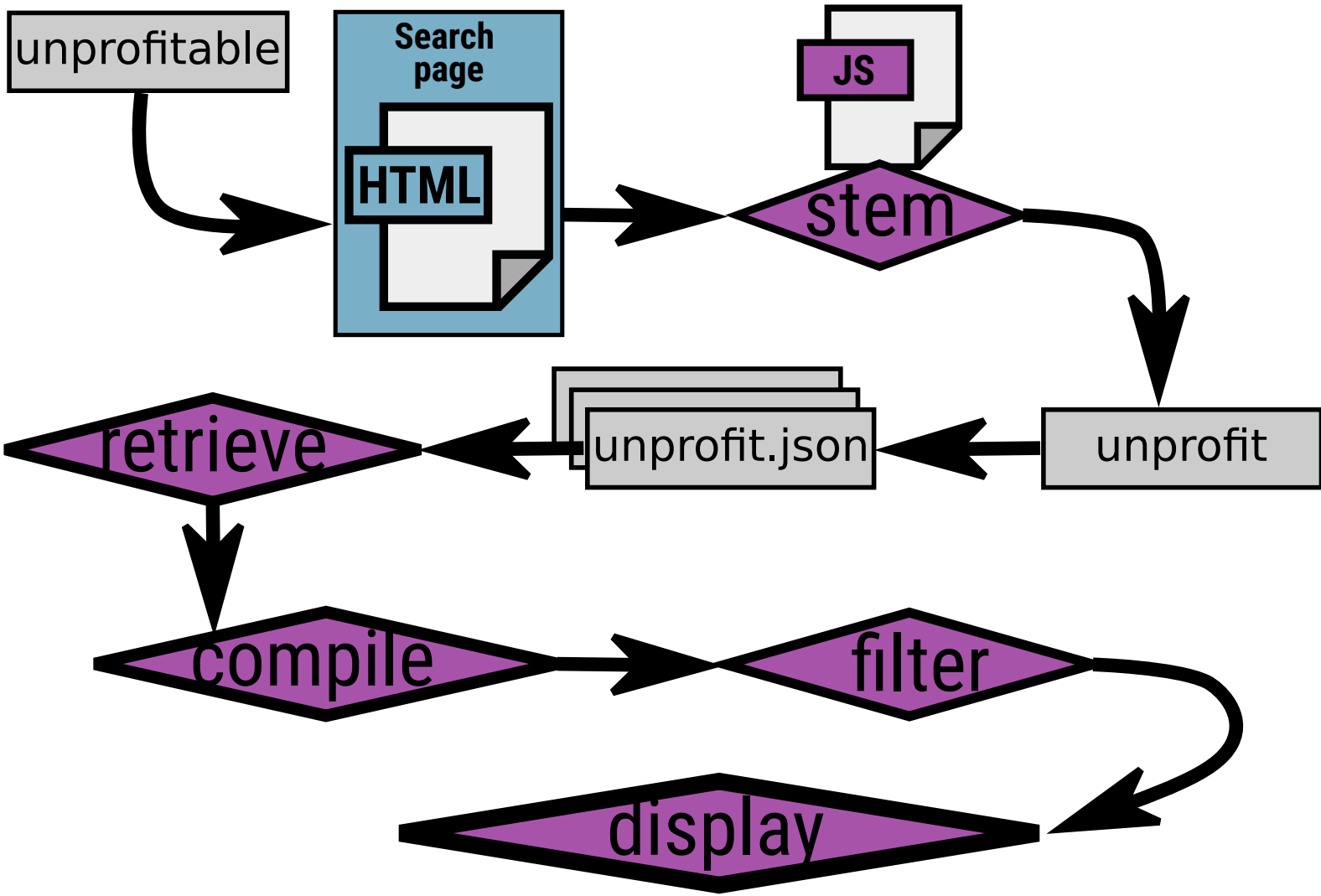












Documents found: 2

I.—The Old Man of Hoy Score: 1

- ...whole that day was spent unprofitably.


John and Joan, Canto II Score: 1

- ...too much ap- propriated unto unprofitable jocularities and facetiousness. Craving licence,...

Endings Projects



= live with staticSearch

-  My Norse Digital Image Repository (myndir.uvic.ca)
-  Map of Early Modern London (mapoflondon.uvic.ca)
-  The Winnifred Eaton Archive (winnifredeatonarchive.org)
-  Mapping Keats's Progress (johnkeats.uvic.ca)
-  Robert Graves Diary (graves.uvic.ca)
- Le mariage sous L'Ancien Régime (mariage.uvic.ca)
- Digital Victorian Periodical Poetry (dvpp.uvic.ca)
- Landscapes of Injustice (loi.uvic.ca)
- Linked Early Modern Drama Online (lemdo.uvic.ca)

Demonstration applications

- <https://johnkeats.uvic.ca/search.html>
- <https://winnifredeatonarchive.org/search.html>
- <https://dvpp.uvic.ca/search.html>

Mappings Keats's Progress



Mapping Keats's Progress

A Critical Chronology

[Home](#)[About](#)[1794-1815](#)[1816](#)[1817](#)[1818](#)[1819](#)[1820](#)[1821](#)[Gallery](#)[Resources](#)[Index](#)

Search this site

If you're searching for proper nouns, capitalize the first letter; otherwise use lower-case. You can search for a quoted phrase, or use plus and minus signs to specify that a term must or must not appear in the result documents. You can also try a [Google Search](#).

Search in

☐ Other pages ☐ Poems ☐ Timeline chapters ☐ Yearly chronology pages

Date range

From:

To:

©2018 [G. Kim Blank](#)

[Comments and suggestions](#) & [Cite this page](#).

Edition 3.3 5th September 2020. SVN revision 1422.



MKP



Search this site

If you're searching for proper nouns, capitalize the first letter; otherwise use lower-case. You can search for a quoted phrase, or use plus and minus signs to specify that a term must or must not appear in the result documents. You can also try a [Google Search](#).



Search in

☐ Other pages ☐ Poems ☐ Timeline chapters ☐ Yearly chronology pages

Date range

From:

To:

Documents found: 82

[11 April 1819: Two Miles & a Thousand Things: A Walking Talk by Coleridge](#) Score: 9

- ...& a Thousand Things: A **Walking** Talk by Coleridge...
- Where, on a **walk** toward Highgate in a...
- ...at the time, and is **walking** with someone Keats knew...
- ...character and sprawling brilliance: I **walked** with him at his...
- ...some of this on his **walk** with Coleridge. (Click image...
- ...& a Thousand Things: A **Walking** Talk by Coleridge "...
- ...& a Thousand Things: A **Walking** Talk by Coleridge ,"
- ...& a Thousand Things: A **Walking** Talk by Coleridge "...

[22 October 1818: Walking With Hazlitt, Thinking Like Hazlitt, Not Playing Racquets Like Hazlitt](#) Score: 6

- 22 October 1818: **Walking** With Hazlitt, Thinking Like...
- Where Keats **walks** with William Hazlitt the

Search this site

If you're searching for proper nouns, capitalize the first letter; otherwise use lower-case. You can search for a quoted phrase, or use plus and minus signs to specify that a term must or must not appear in the result documents. You can also try a [Google Search](#).

Search in

☐ Other pages ☐ Poems ☐ Timeline chapters ☐ Yearly chronology pages

Date range

From:

To:

Documents found: 49

[11 April 1819: Two Miles & a Thousand Things: A Walking Talk by Coleridge](#) Score: 5

- ...& a Thousand Things: A **Walking** Talk by Coleridge...
- ...at the time, and is **walking** with someone Keats knew...
- ...& a Thousand Things: A **Walking** Talk by Coleridge "...
- ...& a Thousand Things: A **Walking** Talk by Coleridge ,"
- ...& a Thousand Things: A **Walking** Talk by Coleridge "...

[22 October 1818: Walking With Hazlitt, Thinking Like Hazlitt, Not Playing Racquets Like Hazlitt](#) Score: 4


- 22 October 1818: **Walking** With Hazlitt, Thinking Like...
- ...G. Kim. "22 October 1818: **Walking** With Hazlitt, Thinking Like...
- ...Kim Blank, "22 October 1818: **Walking** With Hazlitt, Thinking Like...
- ...G. Kim. "22 October 1818: **Walking** With Hazlitt, Thinking Like...

[People mentioned](#) Score: 4

- ...Keats, most famously on their **walking** tour in the summer...
- ...by the Romantics; Keats's northern **walking** tour with Charles Brown...

Search this site

If you're searching for proper nouns, capitalize the first letter; otherwise use lower-case. You can search for a quoted phrase, or use plus and minus signs to specify that a term must or must not appear in the result documents. You can also try a [Google Search](#).



Search in

☐ Other pages ☒ Poems ☐ Timeline chapters ☐ Yearly chronology pages

Date range

From: 1794-01-01

To: 1821-12-31

Documents found: 2

[Endymion: A Poetic Romance BOOK III](#) Score: 1


- He spake, and walking to that aged form,...

[I stood tip-toe upon a little hill](#) Score: 1

- Walking upon the white clouds...

Search this site

If you're searching for proper nouns, capitalize the first letter; otherwise use lower-case. You can search for a quoted phrase, or use plus and minus signs to specify that a term must or must not appear in the result documents. You can also try a [Google Search](#).

<input type="text" value="walking"/>		Search
<input type="button" value="Clear"/>		

Search in

☐ Other pages ☐ Poems ☐ Timeline chapters ☐ Yearly chronology pages

Date range

From:	<input type="text" value="1820"/>
To:	<input type="text" value="1821-12-31"/>
<input type="button" value="Search"/>	

Documents found: 1

- [23 June 1820: Keats Coming Full Circle: Under Hunt's Care](#) Score: 1
- ...set off by an exhausting **walking** tour, August 1818, which...

The Winnifred Eaton Archive

SEARCH FILTERS

Filter

PUBLICATION DATE >

GENRE >

PSEUDONYM >

DOCUMENT TYPE v

☐ Born Digital
document☐ Listing Pages☒ Primary Source:
Manuscripts☒ Primary Source:
Published

EXHIBIT >

garden -Japan -Japanese



Clear

► Instructions

Sort by Relevance (Highest to Lowest) v

Documents found: 3



Marion (Part 1)

► More Info

Next to our garden, separated only by a wooden fence, through...

...used to peep, was the opulent and well-kept garden of Monsieur Prefontaine, who was a very...

We had an old, straggly garden. Everything about it looked "seedy" and uncared...

...and wild, for we could not afford a gardener. My sisters and I found small consolation...

...natural wild state. I was sensitive about that garden. It was awfully poor looking, in comparison...

...said it. Being just next to the fine garden of Monsieur Prefontaine, it appeared the more...

...it appeared the more ragged and beggarly, that garden of ours.

...big, luscious berries I knew were in the garden of Monsieur Prefontaine.

...I kept thinking of the berries in the garden adjoining, and the more I thought of...

It was very quiet in our garden. Not a sound was anywhere, except the...

+1



Plain Pig

► More Info

...route through the kitchen, and out through the garden and the barnyard to the corral, where...

SEARCH FILTERS

Filter

PUBLICATION DATE ▾

From:

1900

To:

1901

GENRE ▸

PSEUDONYM ▸

DOCUMENT TYPE ▾

☐ Born Digital
document☐ Listing Pages☒ Primary Source:
Manuscripts☒ Primary Source:
Published

EXHIBIT ▸

fus[au]ma



Clear

► Instructions

Sort by Relevance (Highest to Lowest) ▾

Documents found: 2



Kirishima-san

▼ More Info

GENRE

Short Story

DOCUMENT TYPE

Primary Source: Published

PUBLICATION DATE

November 1901

EXHIBIT

Playing Japanese 1896-1922

PSEUDONYM

Onoto Watanna

...floor in the next room. Jack pushed the fusuma aside. Kirishima was kneeling on the ground,...

...making some remark about the oddity of the fusuma, but Jack was persistent in a dogged...



The Old Jinrikisha (Part 5)

▼ More Info

GENRE

Novel: Serial

EXHIBIT

Playing Japanese 1896-1922

PUBLICATION DATE

June 1900

DOCUMENT TYPE

Primary Source: Published

PSEUDONYM

Onoto Watanna

...short, and stared with fascinated eyes toward the fusuma which was slowly sliding apart.

SEARCH FILTERS

Filter

PUBLICATION DATE >

GENRE >

PSEUDONYM >

DOCUMENT TYPE >

EXHIBIT >

sam*r*

Clear



► Instructions

Sort by Relevance (Highest to Lowest) ▼

Documents found: 32



The Wooing of Wistaria

▼ More Info

GENRE

Novel

EXHIBIT

Playing Japanese 1896-1922

PSEUDONYM

Onoto Watanna

PUBLICATION DATE

1902

...countless retainers and servants, besides a host of samurai, petty vassals, soldiers, peasants, and citizens, who...

...stone. Its gates were guarded by handsome, bold samurai clad in thick armor. The steel upon...

...attended a private school, kept by an old samurai, where with half a dozen other little...

Sir Genji, the big samurai, who had especial charge of Wistaria's train,...

"Perhaps," admitted the astute samurai, frowning at the same time upon a...

...of them had gone ahead with her father's samurai. Consequently the beat of a horse's hoofs...

...gloomy reflections and uncertainty of purpose, an enormous samurai touched him sharply upon the arm. In...

...of rice-paper protruding from his sleeve. Stealthily the samurai reached out his arm to the lover....

...save that of the sword, when their respective samurai and vassals, coming in contact with one...

...guardian (she being an orphan) was a young samurai in the service of the Prince Mori...

SEARCH FILTERS

Filter

PUBLICATION DATE >

GENRE >

PSEUDONYM ▾

- ☐ The Author Of Me
- ☐ Ellis Eaton
- ☐ Kitishimi Tasha Hasche
- ☒ Mrs. Francis Reeve
- ☒ Mrs. Reeve
- ☒ Mrs. Reeves
- ☐ O. W
- ☐ Onoto Watanna
- ☐ Onoto Watanna
- ☒ Winifred Reeve
- ☒ Winnie Eaton
- ☒ Winnifred Eaton
- ☒ Winnifred Eaton Reeve
- ☒ Winnifred Reeve

DOCUMENT TYPE >

EXHIBIT >

sam*r*



Clear

► Instructions

Sort by Relevance (Highest to Lowest) ▾

Documents found: 2



Honorable Movie Takee Sojin

▼ More Info

DOCUMENT TYPE
Primary Source: Published

GENRE
Non-fiction

EXHIBIT
In Hollywood 1916–1935

PSEUDONYM
Onoto Watanna
Winifred Reeve

PUBLICATION DATE
March 1928

...her husband's ancestry and education. His family were samurai under the Princes of Date. His father...



What Happened to Hayakawa

▼ More Info

DOCUMENT TYPE
Primary Source: Published

GENRE
Non-fiction

EXHIBIT
In Hollywood 1916–1935

PSEUDONYM
Winnifred Eaton Reeve
Onoto Watanna

PUBLICATION DATE

Digital Victorian Periodical Poetry

- Multiple search pages/indexes
- Different purposes
- Different configurations and filters

BETA

DIGITAL VICTORIAN PERIODICAL POETRY

"much to do with victorian poetry"

[Search](#)[Digital edition](#)[Digital index](#)[Poem database](#)[Personography](#)[Diagnostics](#)[Project documentation](#)[Text analysis](#)[About us](#)[Search this site](#)[Search for people](#)[Search figures](#)[Search poem notes](#)

Search for poems based on their figures and illustrations.

If you're searching for proper nouns, capitalize the first letter; otherwise use lower-case. You can search for a quoted phrase, or use plus and minus signs to specify that a term must or must not appear in the result documents.

[Search](#)[Clear](#)

Components of illustrative figures

- ☐ Actions
- ☐ Adult
- ☐ Angel
- ☐ Animal
- ☐ Artwork
- ☐ Baby

Illustration type and placement

- ☐ Caption to illustration is illustrated
- ☐ Decorated or illustrated capital letter
- ☐ Enlarged initial contains a human or animal figure
- ☐ Enlarged initial that contains an illustration
- ☐ Illustrated (decorated) border
- ☐ Illustration above

Periodical

- ☐ Atalanta

Other features

Pluggable stemmers (currently we have only Modern English and the “identity” stemmer).

Next steps

- A Modern French stemmer
- Enhanced English and French stemmers for Early Modern texts
- Option to strip diacritics before stemming

Dev time amortization

- 1 year of work
- 2 contributors, 2 institutions
- Approx 11,000 lines of code
- Nine projects supported and counting
- Value-for-money increases with every project

Takeaways

- A DH Centre like HCMC massively multiplies the value of developer time
- Developer collaborations between institutions add more value
- Minimize custom project code, maximize reusable code

Resources

- Get the code:
<https://github.com/projectEndings/staticSearch>
- Read the documentation:
<https://endings.uvic.ca/staticSearch/docs/>

Thanks!

- HCMC, University of Victoria
- DHIL, Simon Fraser University
- Social Sciences and Humanities Research Council